

## BAB IV

### PENUTUP

#### 4.1. Simpulan

Penelitian ini membahas perbandingan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen menggunakan data yang diberi label secara ahli bahasa (*expert*) dan menggunakan model *lexicon*. Pada proses *Pre-processing* melibatkan *Case Folding*, *Cleaning*, *Tokenizing*, *Stopword Removal*, dan *Stemming* terbukti efektif dalam mempersiapkan data teks untuk analisis. Penggunaan metode TF-IDF dalam pembobotan kata membantu mengidentifikasi kata-kata kunci yang relevan dari setiap ulasan, memungkinkan model untuk lebih memahami konteks dan makna di balik teks.

Hasil penelitian menunjukkan bahwa metode KNN dan Naive Bayes memiliki akurasi yang berbeda berdasarkan sumber label data. Pada data yang diberi label model *lexicon*, akurasi KNN mencapai 78% dan Naive Bayes 65%. Namun, ketika data diberi label oleh *expert*, akurasi kedua metode meningkat menjadi 96%. Perbedaan ini disebabkan oleh perubahan kelas sentimen saat pelabelan menggunakan model *lexicon* dibandingkan dengan pelabelan oleh *expert*, dimana data positif dan netral menurun tetapi pada data negatif mengalami peningkatan. Penelitian ini menyimpulkan bahwa kualitas pelabelan data sangat mempengaruhi akurasi metode klasifikasi yang digunakan, dan pelabelan oleh ahli bahasa (*expert*) dapat meningkatkan keakuratan dalam klasifikasi sentimen.

Selain itu, penelitian ini juga meneliti pengaruh rasio pembagian data dan nilai K pada kinerja model KNN. Hasilnya menunjukkan bahwa rasio pembagian data 70:30 memberikan akurasi terbaik sebesar 96%, dibandingkan dengan rasio 90:10 dan 80:20

yang masing-masing mencapai akurasi 95%. Dalam hal penggunaan nilai K, nilai K1 mencapai akurasi 95.24%, sementara nilai K2 hingga K20 mencapai akurasi yang sama yaitu 96.19%. Berdasarkan hasil ini, nilai K=5 dipilih sebagai parameter final karena memberikan kinerja optimal tanpa penurunan akurasi yang signifikan.

Penelitian ini menyimpulkan, bahwa pentingnya pemilihan metode, rasio data, dan parameter model yang tepat dalam meningkatkan akurasi klasifikasi sentimen. penggunaan pelabelan oleh *expert* dapat meningkatkan dan berkontribusi signifikan terhadap hasil yang diperoleh. dan memberikan informasi tentang efisiensi dan efektivitas dalam memilih metode klasifikasi yang lebih efektif dalam melakukan analisis sentimen yang sesuai dengan kebutuhan penelitian.

#### **4.2.Implikasi**

Penelitian ini memiliki beberapa implikasi penting. Pertama, penggunaan pelabelan data oleh ahli bahasa terbukti meningkatkan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen, sehingga peneliti lain yang menggunakan analisis sentimen dapat mempertimbangkan pelibatan ahli bahasa untuk meningkatkan hasil klasifikasi. Hasil analisis sentimen pada ulasan pengguna aplikasi DANA memberikan wawasan mendalam mengenai keluhan dan masalah yang dihadapi pengguna, yang dapat digunakan untuk meningkatkan fitur dan layanan, sehingga meningkatkan kepuasan pengguna. Penelitian ini juga memberikan perbandingan antara metode KNN dan Naive Bayes dalam mengklasifikasikan sentimen, yang dapat menjadi referensi bagi peneliti lain dalam memilih metode yang sesuai untuk analisis sentimen pada dataset serupa. Penggunaan pembobotan kata dengan metode TF-IDF terbukti efektif dalam proses klasifikasi sentimen, menunjukkan bahwa metode ini dapat diterapkan dalam penelitian lain yang memerlukan analisis teks dengan akurasi tinggi. Akhirnya, pemahaman terhadap sentimen pengguna aplikasi DANA memungkinkan

masyarakat membuat keputusan yang lebih baik dalam memilih aplikasi pembayaran elektronik yang sesuai dengan kebutuhan mereka.